

University of Groningen

Statistical methods for studying the evolution of networks and behavior

Schweinberger, Michael

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version

Publisher's PDF, also known as Version of record

Publication date:

2007

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Schweinberger, M. (2007). *Statistical methods for studying the evolution of networks and behavior*. [Thesis fully internal (DIV), Rijksuniversiteit Groningen]. [S.n.].

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Chapter 6

Discussion

Studying longitudinal network and behavior data is important for understanding social processes, because human beings are interrelated, and the relationships among human beings (human networks) on one hand and human behavior on the other hand are not independent. Examples of dependencies in networks are reciprocity and transitivity (see Wasserman and Faust, 1994). Examples of dependencies between networks and behavior are selection and influence processes. Selection processes operate when actors (e.g., adolescents) change relationships (e.g., friendships), and when selecting relation partners actors take the behavior of potential relation partners (e.g., the consumption of alcohol and drugs) into consideration. Influence processes operate when actors (e.g., adolescents) change behavior (e.g., the consumption of alcohol and drugs), and when changing behavior actors are influenced by the behavior of existing relation partners (e.g., friends). Thus, changes in the network may be affected by behavior, and changes in the behavior may be affected by the network.

As a result of the dependencies, longitudinal network and behavior data are complex. Studying longitudinal network and behavior data therefore demands statistical methods, replacing (complex) data by (simple) summaries of the data (statistics, such as estimates, test statistics) containing as much as possible information about the social processes of interest (cf. Fisher, 1922). Statistical inference for longitudinal network data was pioneered by Holland and Leinhardt (1977) and Snijders (1996, 2001), and extended to longitudinal network and behavior data by Snijders, Steglich, and Schweinberger (2006). These models are based on the assumption that the network and behavior evolution is governed by a Markov process which operates in continuous time but is observed at two or more discrete time points. The Markov process is modeled as driven by actors who, at stochastic times, are allowed to change

either relationships or behavior; and when changing relationships, actors may take into account the behavior of potential relation partners, and when changing behavior, actors may take into account the behavior of existing relation partners. Snijders, Steglich, and Schweinberger (2006) proposed to estimate such models by the method of moments (MM).

The framework of Snijders, Steglich, and Schweinberger (2006) was the point of departure of the present thesis and was advanced by the present thesis both in terms of modeling and statistical inference. The advances made in the present thesis are summed up in Section 6.1. Possible directions of future research are sketched in Section 6.2.

6.1 Summary

The summary of the thesis proceeds chapter by chapter.

Chapter 2: Estimating functions: derivative estimation. In the MM framework of Snijders, Steglich, and Schweinberger (2006), the derivatives of the estimating function with respect to the parameters are required for (1) the standard errors of the parameter estimates; (2) goodness-of-fit test statistics; (3) sensitivity analyses; and (4) linear extrapolation. In the absence of closed-form expressions, the derivatives are estimated by Monte Carlo (MC) methods. Examining the properties of the conventional MC estimator of the derivatives, the finite differences (FD) estimator, revealed that the FD estimator is biased, inconsistent, associated with a bias-variance dilemma (cf. Lemma 1), and expensive in terms of computation time when the number of parameters is moderate or large. Using Lemma 2, three alternative MC estimators were proposed based on the likelihood ratio / score function method of derivative estimation, using variance reduction methods based on control variates. The proposed estimators proved to be consistent and unbiased (see Lemma 2), two of them tend to be more efficient than the FD estimator in terms of computation time, and one of them has minimum variance in a large class of estimators (see Lemma 3). Theoretical insight and two Monte Carlo simulation studies suggested that the FD estimator can bias conclusions considerably, and recommended the use of two of the proposed estimators as alternatives.

Chapter 3: Tests of goodness-of-fit. In the MM framework of Snijders, Steglich, and Schweinberger (2006), goodness-of-fit tests had not been considered before. A test

statistic was proposed that can be regarded as a generalized score test statistic based on regular estimating functions and that admits to test the goodness-of-fit of models which are restricted in one or more parameters. To evaluate the goodness-of-fit test statistic, it is not required to estimate the restricted parameters, which saves computation time and allows to test hard-to-estimate parameters; it was shown that, when the goodness-of-fit of the restricted model is found to be unacceptable, the estimates of all parameters (without restrictions) can be approximated by simplistic one-step estimates, which are crude approximations of the unrestricted estimates, but—other than the unrestricted estimates—require almost no additional computation time. A Monte Carlo simulation study indicated that the finite-sample null distributions of one- and multi-parameter goodness-of-fit tests are close to the expected null distributions; in addition, one-parameter goodness-of-fit tests were compared to conventional one-parameter t -tests. A large, empirical data set was studied and the value of the goodness-of-fit test statistic in forward model selection procedures was demonstrated; the one-parameter goodness-of-fit tests and the one-parameter t -tests turned out to agree by and large; and the one-step estimates appeared to be reasonable approximations of the unrestricted estimates.

Chapter 4: Random effects models. Snijders, Steglich, and Schweinberger (2006) assume that all relevant knowledge with respect to actors is observed in the form of covariates and correctly incorporated in the model. For longitudinal network data, models were proposed that take unobserved heterogeneity across actors into account by means of latent, actor-dependent variables (or random effects), which are governed by a probability law that is common to all actors. Estimating such models involves estimating the variance-covariance matrix of the random effects subject to symmetry and positive definiteness constraints. Maximum likelihood (ML) and Bayesian estimation methods were elaborated; for ML estimation, the non-redundant elements of the random effects variance-covariance matrix were reparametrized so that estimates of the variance-covariance matrix are by construction symmetric and positive definite, and the estimation of very small variances is facilitated. ML and Bayesian estimation were implemented by Markov chain Monte Carlo (MCMC) methods. The random effects models and methods were illustrated by an application to an empirical data set, where ML and Bayesian estimates of the random effects variance-covariance matrix were compared, and, in the Bayesian framework, the sensitivity of the posterior distribution of the random effects variance-covariance matrix to the prior distribution was examined.

Chapter 5: Bayesian modeling and estimation. Network and behavior models were considered, including extensions of the latent variable model of Chapter 4 to longitudinal network and behavior data. The importance of Bayesian inference was stressed, because it (1) is well-suited to studying unique, non-repeatable social processes, and (2) has practical advantages, since when ML algorithms do not converge, the incorporation of prior knowledge can enable Bayesian estimation, and when ML algorithms do converge, then Bayesian estimation can complement ML estimation by giving—under vague prior distributions—insight into the shape of the likelihood function and non-linear dependencies between parameters. A MCMC algorithm for sampling from the posterior distribution was proposed. MM, ML, and Bayesian estimation were compared on the basis of two small, empirical data sets: one data set where MM and ML estimation turned out to be problematic but Bayesian estimation was possible, and another data set where Bayesian and ML estimation agreed by and large, while MM estimation gave rise to deviating results.

Concluding remarks. The statistical methods of Chapters 2 and 3 have already proved to be useful in a number of applications (see, e.g., Knecht, in preparation, and Mercken, in preparation). The experience with respect to the statistical models and methods of Chapters 4 and 5 is limited, and more applications are required to explore the full potential of these models and methods; in the important special case of network models (with exogeneous behavior), however, the Bayesian methods of Chapter 5 have been applied by the author to a large number of simulated and empirical data sets, and have proved to be useful alternatives to MM and ML estimation.

All proposed models and methods were implemented by the author in the computer program **Siena** (Snijders, Steglich, Schweinberger, and Huisman, 2006), which is part of the program collection **StoCNET** (Boer, Huisman, Snijders, Steglich, Wichers, and Zeggelink, 2006) and can be obtained from the website <http://stat.gamma.rug.nl/stocnet>.

6.2 Future research

The present thesis, and in particular Chapters 4 and 5, have opened the gates to a large number of model extensions and applications. First and foremost, the introduction of latent variables and the corresponding estimation framework allows to build and estimate models which make less restrictive assumptions about the data-

generating process than Snijders, Steglich, and Schweinberger (2006), and thus are expected to improve the goodness-of-fit of models. A great number of other latent variable models are conceivable, which remove one or another restrictive assumption about the data-generating process, such as measurement models, where relationships are allowed to be measured with random errors, or multilevel models, where network and behavior are measured on multiple, non-overlapping sets of actors and parameters are allowed to vary within and between sets of actors; both classes of models could be obtained as extensions of the latent variable framework considered here. Other examples of latent variable models are latent structure models, which assume that there is some latent structure that has an impact on the network and behavior evolution, e.g., latent classes to which actors belong and which influence the probability of changes, or latent metric spaces in which actors are positioned (cf. Schweinberger and Snijders, 2003) and which influence the probability of changes through the distances between actors.

Two remarks are in place, however. First, before considering other latent variable models, some issues need to be addressed for the latent variable models proposed here, in the first place, model selection issues (cf. Chapter 4). Second, since the Markov process, which is assumed to govern the network and behavior evolution, is not observed in continuous time but at discrete time points, one is well-advised to keep the model as parsimonious as possible and avoid the use of latent variables unless there are good reasons.

